Contemporary Mathematics

On the gap probability generating function at the spectrum edge in the case of orthogonal symmetry

P.J. Forrester

This paper is dedicated to Percy Deift on the occasion of his 60th birthday

ABSTRACT. The gap probability generating function has as its coefficients the probability of an interval containing exactly k eigenvalues. For scaled random matrices with orthogonal symmetry, and the interval at the hard or soft spectrum edge, the gap probability generating functions have the special property that they can be evaluated in terms of Painlevé transcendents. The derivation of these results makes use of formulas for the same generating function in certain scaled, superimposed ensembles expressed in terms of its correlation functions. It is shown that by a judicious choice of the superimposed ensembles, the scaled limit necessary to derive these formulas can be rigorously justified by a straight forward analysis.

1. Introduction

1.1. An applied setting for gap probabilities. The first use of random matrices to problems in theoretical physics was in relation to the study of the spectra of heavy nuclei (see [**Por65**] for a collection of early works on the subject). It was hypothesized that the highly excited energy levels of heavy nuclei would have the same statistical properties as the the eigenvalues from an ensemble of large random real symmetric matrices. More explicitly, the large random real symmetric matrices were chosen from the Gaussian orthogonal ensemble (GOE) in which each matrix X occurs with probability density $e^{-\mathrm{Tr}X^2/2}$ (such an ensemble is invariant under the transformation $X \mapsto OXO^T$ where O is a real orthogonal matrix; this has the physical interpretation of there being no preferential basis and explains too the adjective orthogonal in GOE). To leading order matrices from the GOE have the support of their eigenvalues in $[-\sqrt{2N}, \sqrt{2N}]$. The largest eigenvalue thus occurs in the neighbourhood of $\sqrt{2N}$, which is referred to as the soft edge, while the region away from the edges (for example in the neighbourhood of the origin) is referred to as the bulk. It is the statistical properties of the eigenvalues of large

2000 *Mathematics Subject Classification.* Primary 15A52 ;Secondary 60K37.
Key words and phrases. Random matrices, spacing distribution.
Supported by the Australian Research Council.

GOE matrices in the bulk, scaled so that their mean spacing is unity, which are compared against data from the spectra of heavy nuclei (with the latter also scaled so that the mean spacing between consecutive levels is unity).

More recently problems from statistical physics have led to applications of distributions at the soft edge of random matrix ensembles (i.e. in the neighbourhood of the largest eigenvalue.) Here each of the three symmetry classes — orthogonal, unitary and symplectic — are relevant to the applications. In terms of Gaussian matrices, such symmetry classes are realized by the probability density $e^{-\beta \mathrm{Tr} X^2/2}$ in which the Hermitian matrix X has real elements in the case of orthogonal symmetry ($\beta = 1$), complex elements in the case of unitary symmetry ($\beta = 2$) and real quaternion elements represented as 2×2 matrices in the case of symplectic symmetry ($\beta = 4$). In the latter case the corresponding matrix X has doubly degenerate eigenvalues, and the convention is to count only one of the distinct eigenvalues in the trace. For each of the three symmetry classes the corresponding eigenvalue probability density function (p.d.f.) is of the form

$$\frac{1}{C}\prod_{l=1}^{N} g_\beta(x_l) \prod_{1\le j<k\le N} |x_k - x_j|^\beta \tag{1.1}$$

with $g(x) = e^{-\beta x^2}$ (C denotes the normalization). Independent of β the leading order support is on $[-\sqrt{2N}, \sqrt{2N}]$, and in the neighbourhood of the largest eigenvalue at $x = \sqrt{2N}$ (the soft edge) the eigenvalues have spacing $O(1/N^{1/6})$. With $E_{N,\beta}(k;(s,\infty);g_\beta(x))$ denoting the probability that exactly k eigenvalues are in the interval (s,∞) of the ensemble as specified by (1.1) ($E_{N,\beta}$ is broadly referred to as a gap probabilty), in keeping with these facts one expects

$$\lim_{N\to\infty} E_{N,\beta}\Big(k;(\sqrt{2N} + c_\beta X/N^{1/6},\infty); e^{-\beta x^2/2}\Big) \tag{1.2}$$

to be a well defined order one quantity (here c_β is an N independent factor chosen for convenience). Significantly, it is also expected that for all $g_\beta(x)$ such that the eigenvalue support is to leading order a single interval with right endpoint $a(N)$, there will be a scale $b(N)$ such that

$$\lim_{N\to\infty} E_{N,\beta}\Big(k;(a(N) + b(N)X,\infty); g_\beta(x)\Big) \tag{1.3}$$

exists and is equal to (1.2). Such universality questions have been a major theme of P. Deift and collaborators (see [**Dei99**] for a summary of this work up to 1999 relating to $\beta = 2$, and [**DG04, DG06, CDG06, DGKV06**] for recent results on $\beta = 1$ and 4). In particular, with

$$g_\beta(x) = x^a e^{-\beta x/2}, \qquad (x > 0)$$

which corresponds to the so called Laguerre ensemble, it is known from the rigorous work of Johansson and Johnstone [**Joh00, Joh01**] that for $\beta = 1$ and 2

$$\begin{aligned}
&\lim_{N\to\infty} E_{N,\beta}(k;(\sqrt{2N} + X/\sqrt{2}N^{1/6},\infty); e^{-\beta x^2}) \\
&= \lim_{N\to\infty} E_{N,\beta}(k;(4N + 2(2N)^{1/3},\infty); x^a e^{-\beta x/2}).
\end{aligned}$$

This limiting probability is denoted $E_\beta^{\mathrm{soft}}(k;(s,\infty))$.

One class of problems in statistical physics giving rise to these probabilities relates to a last passage percolation problem, originally formulated by Hammersley,

and various symmerizations due to Baik and Rains [**BR01**]. Here a unit square, with bottom left corner at the origin, contains points uniformly at random with a Poisson rate of intensity z^2 (this means that for δ small, each non-overlapping $\delta \times \delta$ subsquare has probability $z^2\delta^2$ of containing a point). Continuous piecewise linear paths are formed from $(0,0)$ to $(1,1)$ by joining points in the square with segments of positive slope. The length of a path is defined as the number of points it passes through, and l^U is used to denote the maximum of the length of all possible paths. The limit theorem of Baik, Deift and Johansson [**BDJ99**] tells us that

$$\lim_{z\to\infty} \Pr\Big(\frac{l^U - 2z}{z^{1/3}} \le y\Big) = E_2^{\rm soft}(0;(y,\infty)),$$

so relating to the soft edge gap probability with $\beta = 2$. In regards to this probability with $\beta = 1$, let $l^S/2$ denote the maximum length of all paths going from $(0,0)$ to this diagonal. A limit theorem of Baik and Rains [**BR01**] gives

$$\lim_{z\to\infty} \Pr\Big(\frac{l^S - 2z}{z^{1/3}} \le y\Big) = E_1^{\rm soft}(0;(y,\infty)).$$

Furthermore, with the points constrained to be symmetrical about the lower left to upper right diagonal, Baik and Rains [**BR01**] proved an analogous limit theorem relating to $E_4^{\rm soft}$.

1.2. The gap probability generating function and Fredholm determinants. A special feature of the gap probabilities $\{E_\beta^{\rm soft}(k;(s,\infty))\}_{k=0,1,\dots}$ for the random matrix couplings $\beta = 1, 2$ and 4 is that they can be expressed in terms of Fredholm determinants and Painlevé transcendents. How this comes about has different features for each of the three β values. However a common step is that one introduces the generating function

$$E_{N,\beta}(J;g(x);\zeta) = \sum_{k=0}^{\infty}(1-\xi)^k E_{N,\beta}(k;J;g_\beta(x))$$

(because $E_{N,\beta}(k;J;g_\beta(x)) = 0$ for $k > N$ the sum terminantes at $k = N$). About $\xi = 0$ this has the expansion

$$E_{N,\beta}(J;g(x);\zeta) = 1 + \sum_{n=1}^{\infty} \frac{(-\xi)^n}{n!} \int_J dx_1 \cdots \int_J dx_n\, \rho_{(n)}(x_1,\dots,x_n) \tag{1.4}$$

where $\rho_{(n)}$ denotes the n-point correlation function for the point process specified by (1.1).

The case $\beta = 2$ is the simplest, because the p.d.f. (1.1) then specifies a determinantal point process, which means that its n-point correlation is an $n \times n$ determinant. Explicitly, with $\{p_j(x)\}_{j=0,1,\dots}$ denoting the monic orthogonal polynomials with respect to the weight function $g_2(x)$, and $(a,b)_2 := \int_{-\infty}^{\infty} g_2(x)a(x)b(x)\,dx$, one has that (see e.g. [**For**])

$$\rho_{(n)}(x_1,\dots,x_n) = \det\Big[K_N(x_i,x_j)\Big]_{i,j=1,\dots,n} \tag{1.5}$$

where

$$K_N(x,y) = (g_2(x)g_2(y))^{1/2} \sum_{j=0}^{N-1} \frac{p_j(x)p_j(y)}{(p_j,p_j)_2}$$
$$= \frac{(g_2(x)g_2(y))^{1/2}}{(p_{N-1},p_{N-1})_2} \frac{p_N(x)p_{N-1}(y) - p_{N-1}(x)p_N(y)}{x-y}, \tag{1.6}$$

with the final equality due to the Christoffel-Darboux summation formula. Substituting (1.5) in (1.4) one recognises the final expression as an expansion of a Fredholm determinant (see [**WW65**]), giving

$$E_{N,2}(J; g_2(x); \xi) = \det(\mathbb{I} - \xi K_{N,J}) \tag{1.7}$$

where $K_{N,J}$ denotes the integral operator on J with kernel (1.6). With $g_2(x)$ corresponding to the Gaussian or Laguerre weights as introduced above, the soft edge scaling limit is easy to perform rigorously [**Joh00**], leading to the result [**For93**]

$$E_{N,2}^{\rm soft}((s,\infty); \xi) = \det(\mathbb{I} - \xi K_{(s,\infty)}^{\rm soft}) \tag{1.8}$$

where $K_{(s,\infty)}^{\rm soft}$ is the integral operator on (s,∞) with kernel

$$K^{\rm soft}(x,y) = \frac{\mathrm{Ai}(x)\mathrm{Ai}'(y) - \mathrm{Ai}(y)\mathrm{Ai}'(x)}{x-y}. \tag{1.9}$$

This is the sought Fredholm determinant evaluation, and it in turn can be used [**TW94, AvM95, BD02**] to deduce a Painlevé transcendent evaluation.

The situation at $\beta = 1$ and 4 is more complex. The correlations are now quaternion determinants rather than scalar determinants. In relation to the former, let X be a $2N \times 2N$ antisymmetric matrix, and set

$$Z_{2N} := \mathbb{I}_N \otimes \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

With the quaternion dual of a $2N \times 2N$ matrix specified by

$$A^D = Z_{2N} A^T Z_{2N}^{-1}$$

one notes that matrices of the form XZ_{2N} are self quaternion dual. The quaternion determinant qdet of such matrices has the key property [**Dys72**] that

$$\mathrm{qdet}\, XZ_{2N} = \mathrm{Pf}\, X$$

where Pf denotes the Pfaffian. As a consequence, for A self quaternion dual, the quaternion determinant and ordinary determinant are related by

$$(\mathrm{qdet}\, A)^2 = \det A. \tag{1.10}$$

Analogous to (1.7), one has that if, as is the case at $\beta = 1$ and 4 (see e.g. [**For**]),

$$\rho_{(n)}(x_1,\dots,x_n) = \det \left[\tilde{K}_N(x_i,x_j) \right]_{i,j=1,\dots,n}$$

where $\tilde{K}_N$ is a 2×2 matrix such that $[\tilde{K}_N(x_i,x_j)]_{i,j=1,\dots,n}$ is self quaternion dual, then

$$E_{N,\beta}(J; g_\beta(x); \xi) = \mathrm{qdet}(\mathbb{I} - \xi \tilde{K}_{N,J}).$$

Here $\tilde{K}_{N,J}$ is a 2×2 matrix integral operator on the interval J, and $\tilde{K}_{N,J}$ has kernel $\tilde{K}_N(x,y)$. Squaring both sides and applying (1.10) shows

$$\Big(E_{N,\beta}(J; g_\beta(x); \xi) \Big)^2 = \det(\mathbb{I} - \xi \tilde{K}_{N,J}). \tag{1.11}$$

The formula (1.11) with $\xi = 1$ is taken as the starting point of the analysis in [**TW96**] leading to Painlevé evaluations of $E_1^{\rm soft}(0;(s,\infty))$ and $E_4^{\rm soft}(0;(s,\infty))$. This strategy is used in [**Die05**] to give the analogous formulas for the corresponding generating functions. However the details of these calculations are very technical. Fortunately there is another approach to the problem introduced by the present author in [**For00**], and to be further developed herein.

In this approach, instead of working with the soft edge scaled limit of (1.11), and thereby involving a matrix integral operator, the first step is to derive the alternative expression

$$\Big(E_1^{\rm soft}(0;(s,\infty))\Big)^2 = E_2^{\rm soft}(0;(s,\infty))\Big(1 - \int_s^\infty (\mathbb{I} - K_{(s,\infty)}^{\rm soft})^{-1} A^{\rm s}[y] B^{\rm s}(y)\,dy\Big) \tag{1.12}$$

where $K_{(s,\infty)}^{\rm soft}$ is as in (1.8), $A^{\rm s}$ is the operator which multiplies by $\mathrm{Ai}(x)$, while $B^{\rm s}$ is the integral operator with kernel $\int_0^\infty \mathrm{Ai}(y-v)\,dv$. Starting from (1.12) the Painlevé expression can be deduced in a page or two of working (assuming knowledge of some identities from [**TW94**]). Another significant feature of (1.12) is that it has been taken as the starting point of the proof of the Fredholm determinant formula [**Sas05, FS05**]

$$E_1^{\rm soft}(0;(s,\infty) = \det(\mathbb{I} - V_{(0,\infty)}^{\rm soft})$$

where $V^{\rm soft}(x,y)$ is the integral operator on $(0,\infty)$ with kernel $V^{\rm soft}(x,y) = \mathrm{Ai}(x+y+s)$.

To extend the approach of [**For00**] to the Painlevé evaluation of the generating function $E_1^{\rm soft}((s,\infty);\xi)$ [**For06**], a key identity is the formula

$$E^{{\rm odd}({\rm OEsoft})^2}((s,\infty);\xi) = \det(\mathbb{I}-\xi K_{(s,\infty)}^{\rm soft})\Big(1-\xi\int_s^\infty [(\mathbb{I}-\xi K_{(s,\infty)}^{\rm soft})^{-1}A^{\rm s}](y) B^{\rm s}(y)\,dy\Big) \tag{1.13}$$

which for $\xi = 1$ is equivalent to (1.12). Here, with the ensemble (1.1) for $\beta = 1$ referred to as $\mathrm{OE}_N(g_1(x))$ (here the "O" denotes the underlying orthogonal symmetry), $\mathrm{OE}_N(g_1(x))\cup\mathrm{OE}_N(g_1(x))$ denoting the superposition of two independent such ensembles, and the operation "odd" referring to observing only each odd labelled eigenvalue in the superposition as counted from the right most eigenvalue (i.e. the soft edge), $\mathrm{odd}(\mathrm{OEsoft})^2$ refers to the soft edge scaling limit of the ensemble

$$\mathrm{odd}\Big(\mathrm{OE}_N(e^{-x^2}) \cup \mathrm{OE}_N(e^{-x^2})\Big). \tag{1.14}$$

Another advantage of the approach of [**For00**] is that the Painlevé evaluations for $\beta = 4$ are deduced as a corollary of those at $\beta = 1$ and $\beta = 2$. This is possible because of inter-relations between the gap probabilities for the three symmetry types [**FR01**].

The primary motivating factor in seeking an alternative approach to the derivation of the Painlevé evaluations of the $\beta = 1$ and 4 soft edge gap probabilities was to calculate analogous formulas for the hard edge. The hard edge refers to the neighbourhood of the origin when $g_\beta(x) \sim x^a$ as $x \to 0^+$, $g_\beta(x) = 0$ for $x < 0$. In the case $g_\beta(x) = x^a e^{-\beta x/2}$, $x > 0$, the appropriate hard edge scaling is $x \mapsto X/4N$, and working based on superimposed ensembles can be carried through, leading to the sought Painlevé evaluations [**For00, For06**]. More explicitly, let

$$\mathrm{odd}\Big(\mathrm{OE}_N(x^{(a-1)/2}e^{-x/2}) \cup \mathrm{OE}_N(x^{(a-1)/2}e^{-x/2})\Big) =: \mathrm{even}(\mathrm{LOE}_N)^2 \tag{1.15}$$

denote the joint distribution of all odd labelled eigenvalues in the superposition (labelled from the hard edge at $x = 0$). Let odd(OEhard)^2 refer to the hard edge scaling of this joint distribution. The identity which plays the role of (1.13) at the hard edge is

$$E^{\mathrm{odd(OEhard)}^2}((0,s);\xi;\alpha)\Big|_{\alpha=(a-1)/2}$$
$$= \det(\mathbb{I} - \xi K^{\mathrm{hard}}_{(0,s)})\Big(1 - \xi \int_s^\infty [(\mathbb{I} - \xi K^{\mathrm{hard}}_{(0,s)})^{-1} A^{\mathrm{h}}](y) B^{\mathrm{h}}(y)\, dy\Big). \tag{1.16}$$

Here the argument α refers to the hard edge singularity x^α, $K^{\mathrm{hard}}_{(0,s)}$ is the integral operator on $(0,s)$ with kernel

$$K^{\mathrm{hard}}(x,y) = \frac{J_a(\sqrt{x})\sqrt{y}J_a'(\sqrt{y}) - \sqrt{x}J_a'(\sqrt{x})J_a(\sqrt{y})}{2(x-y)} \tag{1.17}$$

while $A^{\mathrm{h}}(x,y)$ is the operator which multiplies by $J_a(\sqrt{x})$ and B^{h} is the integral operator on $(0,s)$ with kernel $\frac{1}{2\sqrt{y}}\int_y^\infty J_a(t)\, dt$.

1.3. Aim of the paper. In this paper we will reconsider the derivation of (1.13) and (1.16), which in the unpublished work [**For00**] is not rigorous. We will begin by recalling and giving a critique of the strategy used in [**For00**]. Then we will proceed to present a rigorous strategy based on different ensembles than those used in (1.13) and (1.16) to scale to odd(OEsoft)^2 and odd(OEhard)^2 respectively. These ensembles have the advantage that the closed form expressions for the correlation functions are simpler than those of the original ensembles. This makes the analysis of the scaling of the corresponding gap probabilities much simpler.

2. Gap probability generating function for superimposed ensembles

2.1. Review and critique of the original calculation. To study the derivation of (1.13) and (1.16) one first notes that for an integral operator $\mathbb{I} + C \otimes D$, the fact that $C \otimes D$ is of rank 1 gives that

$$\det(\mathbb{I} + C \otimes D) = 1 + \int_{-\infty}^\infty C(y) D(y)\, dy.$$

It follows that (1.13) and (1.16) can be rewritten

$$E^{\mathrm{odd(OEsoft)}^2}((s,\infty);\xi) = \det\Big(\mathbb{I} - \xi\Big(\tilde{K}^{\mathrm{soft}}_{(s,\infty)}) + \tilde{A}^{\mathrm{s}} \otimes \tilde{B}^{\mathrm{s}}\Big)\Big) \tag{2.1}$$

$$E^{\mathrm{odd(OEhard)}^2}((0,s);\xi;(a-1)/2) = \det\Big(\mathbb{I} - \xi\Big(\tilde{K}^{\mathrm{hard}}_{(0,s)}) + \tilde{A}^{\mathrm{h}} \otimes \tilde{B}^{\mathrm{h}}\Big)\Big). \tag{2.2}$$

Here the tilde on the operators indicates that the kernels have been multiplied by the gauge factor $(1/\tilde{a}(x))\tilde{a}(y)$ for $\tilde{a}$ decaying sufficiently rapidly so that the integral operators inside the determinants are trace class. This latter technicality can be avoided by expanding (2.1) and (2.2) according to the right hand side of (1.4) to obtain

$$E^{\mathrm{odd(OEsoft)}^2}((s,\infty);\xi)$$
$$= 1 + \sum_{n=1}^\infty \frac{(-\xi)^n}{n!} \int_s^\infty dx_1 \cdots \int_s^\infty dx_n\, \rho^{\mathrm{odd(OEsoft)}^2}_{(n)}(x_1,\ldots,x_n) \tag{2.3}$$

with

$$\rho_{(n)}^{\mathrm{odd}(\mathrm{OEsoft})^2}(x_1,\dots,x_n) = \det\Big[K^{\mathrm{soft}}(x_j,x_k) + \mathrm{Ai}(x_j)\int_0^\infty \mathrm{Ai}(x_k - v)\,dv\Big]_{j,k=1,\dots,n} \tag{2.4}$$

and

$$\begin{aligned} &E^{\mathrm{odd}(\mathrm{OEhard})^2}((0,s);\xi;(a-1)/2) \\ &\quad = 1 + \sum_{n=1}^\infty \frac{(-\xi)^n}{n!}\int_0^s dx_1\cdots\int_0^s dx_n\,\rho_{(n)}^{\mathrm{odd}(\mathrm{OEhard})^2}(x_1,\dots,x_n) \end{aligned} \tag{2.5}$$

with

$$\rho_{(n)}^{\mathrm{odd}(\mathrm{OEhard})^2}(x_1,\dots,x_n) = \det\Big[K^{\mathrm{hard}}(x_j,x_k) + \frac{J_a(\sqrt{x_j})}{2\sqrt{x_k}}\int_{\sqrt{x_k}}^\infty J_a(t)\,dt\Big]_{j,k=1,\dots,n}. \tag{2.6}$$

The derivation of (2.3) given in [**For00**] took as its starting point an explicit form of the n-point correlations

$$\begin{aligned} \rho_{(n)}^{\mathrm{odd}(\mathrm{GOE}_N)^2}(x_1,\dots,x_n) &= \det\Big[K_{N-1}(x_j,x_k) \\ &\quad + e^{-x_j^2/2}2^{-(N-1)}H_{N-1}(x_j)\Big(A_1(x_k)+A_2(x_k)\Big)\Big]_{j,k=1,\dots,n} \end{aligned} \tag{2.7}$$

where K_{N-1} is specified by (1.6) with $g_2(x) = e^{-x^2}$, and

$$\begin{aligned} A_1(y) &= \frac{e^{-y^2/2}}{(N/2-1)!}\sum_{\nu=0}^\infty \frac{(N/2-1+\nu)!}{(N-1+2\nu)!}H_{N-1+2\nu}(y) \\ A_2(y) &= \frac{\pi^{1/2}e^{-y^2/2}}{(N/2-1)!}\sum_{l=0}^\infty \frac{1}{2^{N+2l}(N/2+l)!}H_{N+2l}(y) \end{aligned} \tag{2.8}$$

(in these formulas it is assumed that N is even). The task now is to show that with $J = (s,\infty)$ and the soft edge scale $s \mapsto \sqrt{2N} + s/2^{1/2}N^{1/6}$, the large N limit of (1.4) with $\rho_{(2)}$ given by (2.7) is equal to (2.3). We know from [**Sos02**] (see [**BF03**] for a restatement of this) that for this task it is sufficient to show

$$\begin{aligned} &\lim_{N\to\infty}\int_{\sqrt{2N}+s/2^{1/2}N^{1/6}}^\infty dx_1\cdots\int_{\sqrt{2N}+s/2^{1/2}N^{1/6}}^\infty dx_n\,\rho_{(n)}^{\mathrm{odd}(\mathrm{GOE}_N)^2}(x_1,\dots,x_n) \\ &\quad = \int_s^\infty dx_1\cdots\int_s^\infty dx_n\,\rho_{(n)}^{\mathrm{odd}(\mathrm{OEsoft})^2}(x_1,\dots,x_n). \end{aligned} \tag{2.9}$$

The starting point of the derivation of (2.5) given in [**For00**] is very similar. In the notation of (1.15), the analogue of (2.7) is

$$\begin{aligned} \rho_{(n)}^{\mathrm{odd}(\mathrm{LOE}_N)^2}(x_1,\dots,x_n) &= \det\Big[K_{N-1}(x_j,x_k) + (g_2(x_j))^{1/2}L_{N-1}^a(x_j) \\ &\quad \times\frac{(N-1)!}{2^{N-2}((N-2)/2)!(a/2+(N-2)/2)!}\Big(B_1(x_k)+B_2(x_k)\Big)\Big]_{j,k=1,\dots,n}. \end{aligned} \tag{2.10}$$

Here $K_{N-1}(x,y)$ refers to (1.6) with $g_2(x) = x^a e^{-x}$ and (assuming N is even)

$$B_1(y) = \sum_{\nu=(N-2)/2}^{\infty} \frac{2^{2\nu}(a/2+\nu)!\nu!}{(a+2\nu+1)!}(g_2(y))^{1/2} L^a_{2\nu+1}(y)$$

$$B_2(y) = 2^{a-1}\frac{((a-1)/2)!^2(a/2)!^2}{a!^2}\sum_{l=N/2}^{\infty}\frac{(2l)!}{2^{2l}l!(a/2+l)!}(g_2(y))^{1/2}L_{2l}(y). \tag{2.11}$$

Here the task is to show that with $J=(0,s)$, and the hard edge scale $s \mapsto s/4N$, the large N limit of (1.4) with $\rho_{(2)}$ given by (2.1) is equal to (2.5). For this, according to [**Sos02**], it is sufficient to show

$$\begin{aligned}&\lim_{N\to\infty}\int_0^{s/4N} dx_1 \cdots \int_0^{s/4N} dx_n\, \rho_{(n)}^{\mathrm{odd}(\mathrm{LOE}_N)^2}(x_1,\dots,x_n)\\ &\quad = \int_0^s dx_1 \cdots \int_0^s dx_n\, \rho_{(n)}^{\mathrm{odd}(\mathrm{OEhard})^2}(x_1,\dots,x_n).\end{aligned} \tag{2.12}$$

A mechanism for the validity of (2.9) and (2.12) is the uniform estimates

$$\begin{aligned}&\Big(\frac{1}{\sqrt{2}N^{1/6}}\Big)^n \rho_{(n)}^{\mathrm{odd}(\mathrm{LOE}_N)^2}(\sqrt{2N}+x_1/2^{1/2}N^{1/6},\dots,\sqrt{2N}+x_n/2^{1/2}N^{1/6})\\ &\quad = \rho_{(n)}^{\mathrm{odd}(\mathrm{OEsoft})^2}(x_1,\dots,x_n) + o(1)R_n^{\mathrm{s}}(x_1,\dots,x_n)\end{aligned} \tag{2.13}$$

and

$$\begin{aligned}&\Big(\frac{1}{4N}\Big)^n \rho_{(n)}^{\mathrm{odd}(\mathrm{LOE}_N)^2}(x_1/4N,\dots,x_n/4N)\\ &\quad = \rho_{(n)}^{\mathrm{odd}(\mathrm{OEhard})^2}(x_1,\dots,x_n) + o(1)R_n^{\mathrm{h}}(x_1,\dots,x_n)\end{aligned} \tag{2.14}$$

where $o(1)$ refers to the dependence on N, while R_n is integrable on the domain in question. In [**For00**] only the leading term in (2.13) and (2.14) was computed, so in particular (2.3) and (2.5) were not rigorously established. This "technical issue", essentially asking for uniform estimates of the infinite sums (2.8) and (2.1), is the reason that [**For00**] was not submitted for publication.

2.2. Special superimposed ensembles. With the task of providing uniform asymptotics of (2.8) and (2.1) being technically difficult, necessity dictates seeking an alternative strategy. For this one should bring to the fore the notion of universality, which tells us (for example) that there is nothing canonical about the finite N ensemble (1.14) in regard to studying the limiting distribution $\mathrm{odd}(\mathrm{OEsoft})^2$. Thus for a general weight function $g_1(x)$

$$\begin{aligned}E_1(n;(s,\infty);\mathrm{odd}(\mathrm{OE}_N(g_1(x))\cup\mathrm{OE}_N(g_1(x))) &= \sum_{l=0}^{2n} E_1(2l-1;(s,\infty);g_1(x))\\ &\times\Big(E_1(l;(s,\infty);g_1(x)) + E_1(l-1;(s,\infty);g_1(x))\Big)\end{aligned} \tag{2.15}$$

which follows immediately from the definition of the superimposed ensemble. Hence, the universality of the gap probability in the superimposed ensemble is a consequence of the universality at the edge of the ensemble $\mathrm{OE}_N(g_1(x))$, which is known from [**DG06**]. In particular, instead of considering the soft edge scaling of the

superposition of Gaussian orthogonal ensembles, we may just as well consider the soft edge scaling of superimposed Laguerre orthogonal ensembles

$$\text{odd}\Big(\mathrm{OE}_N(x^a e^{-x/2}) \cup \mathrm{OE}_N(x^a e^{-x/2})\Big), \tag{2.16}$$

and more particularly this for any one value of a. So the question now is, amongst the ensembles (2.16) is there a value of the parameter a for which the correlations have an explicit form which is easier to analyze that that in (2.7)?

That there is a special ensemble amongst (2.16) is seen from a theorem in [**FR01**]. This theorem classifies all continuous weight functions g_1, g_2 such that

$$\text{even}\Big(\mathrm{OE}_n(g_1) \cup \mathrm{OE}_n(g_1)\Big) = \mathrm{UE}_n(g_2) \tag{2.17}$$

where $\mathrm{UE}_n(g_2)$ refers to (1.1) with $\beta = 2$ (here the U denotes unitary symmetry and even refers to the labelling of the eigenvalues countered from the right, which is the soft edge). In fact up to a fractional linear transformation there are only two weight functions with this property,

$$(g_1, g_2) = \begin{cases} (e^{-x/2}, e^{-x}), & x > 0 \\ ((1-x)^{(a-1)/2}, (1-x)^a), & -1 < x < 1. \end{cases} \tag{2.18}$$

Hence of the superimposed Laguerre ensembles in (2.16), the case $a = 0$ is distinguished by the property (2.17). In keeping with this special feature, for the correlations of every odd labelled eigenvalue as required by (2.16), it allows the structured formula [**FR04**]

$$\rho_{(n)}^{\mathrm{odd}(\mathrm{LOE}_N^0)^2}(x_1, \dots, x_n) = \det\Big[-\frac{\partial}{\partial x_j} \int_0^{x_k} K_N(x_j, u)\, du \Big]_{j,k=1,\dots,n} \tag{2.19}$$

where K_N refers to (1.6) with $g_2(x) = e^{-x}$, $p_j(x) = L_j^0(x)$ (here $L_j^0(x)$ denotes the Laguerre polynomials with parameter $a = 0$).

One sees immediately that the structure exhibited in (2.19) gives a much cleaner expression than that exhibited by (2.7). The task is to compute uniform asymptotics of this under soft edge scaling, which for the $a = 0$ Laguerre ensemble is obtained by replacing coordinates [**For93**] $x \mapsto 4N + 2(2N)^{1/3} X$ then taking the limit $N \to \infty$. It is a straightforward exercise using the uniform estimate [**Olv74**]

$$e^{-x/2} L_N^0(x) = \frac{(-1)^N}{(2N)^{1/3}} \mathrm{Ai}(t) + O(e^{-t}) o(N^{-1/3}) \tag{2.20}$$

where $x = 4N + 2(2N)^{1/3} t$, valid for $t \in [t_0, \infty)$ to obtain the uniform asymptotic expansion

$$2(2N)^{1/3} K_N(4N + (2N)^{1/3} s, 4N + (2N)^{1/3} t) = K^{\mathrm{soft}}(s,t) + O(e^{-(s+t)}) O(N^{-1/3}) \tag{2.21}$$

valid for $t, s \in [t_0, \infty)$. Further, this expansion can be differentiated with respect to t or s. However, as written in (2.19) the argument u in $K_N(x_j, u)$ takes on values which are to leading order in $[0, 4N]$ instead of $[4N, \infty)$ as in (2.21). As noted in [**FR04**, Lemma 13], this can be circumvented by using the identity

$$\int_0^x K_N(y,u)\, dy = (-1)^{N-1} \int_y^\infty e^{-u/2} \frac{d}{du} L_N^0(u)\, du - \int_x^\infty K_N(y,u)\, du.$$

Use of (2.20) and (2.21) then gives the uniform asymptotic expansion

$$
\begin{aligned}
&2(2N)^{1/3} \int_0^{4N+2(2N)^{1/3}X} K_N(4N+2(2N)^{1/3}Y, u)\, du \\
&\quad = \int_Y^\infty \mathrm{Ai}(u)\, du - \int_X^\infty K^{\rm soft}(Y,u)\, du + O(e^{-(X+Y)})O(N^{-1/3}),
\end{aligned} \tag{2.22}
$$

which furthermore remains valid upon differentiating with respect to X or Y. Noting from the integral form of the kernel (1.9),

$$
K^{\rm soft}(x,y) = \int_0^\infty \mathrm{Ai}(x+u)\mathrm{Ai}(y+u)\, du
$$

that

$$
-\frac{\partial}{\partial Y} \int_{-\infty}^X K^{\rm soft}(Y,u)\, du = K^{\rm soft}(X,Y) + \mathrm{Ai}(Y) \int_{-\infty}^X \mathrm{Ai}(t)\, dt,
$$

and substituting this in (2.21) then substituting the result in (2.19) we deduce

$$
\begin{aligned}
&(2(2N)^{1/3})^n \rho_{(n)}^{\mathrm{odd}(\mathrm{LOE}_N^0)^2}(4N+2(2N)^{1/3}x_1, \dots, 4N+2(2N)^{1/3}x_N) \\
&\quad = \rho_{(n)}^{\mathrm{odd}(\mathrm{OEsoft})^2}(x_1,\dots,x_n) + e^{-(x_1+\cdots+x_n)}O(N^{-1/3})
\end{aligned} \tag{2.23}
$$

where $\rho_{(n)}^{\mathrm{odd}(\mathrm{OEsoft})^2}$ is specified by (2.4). It is immediate from this that the analogue of (2.9) holds true. But according to [**Sos02**] the latter is sufficient for the validity of the limit formula

$$
\lim_{N\to\infty} E_N^{\mathrm{odd}(\mathrm{LOE}_N^0)^2}((4N+2(2N)^{1/3}s,\infty);\xi) = E^{\mathrm{odd}(\mathrm{OEsoft})^2}((s,\infty);\xi), \tag{2.24}
$$

with $E^{\mathrm{odd}(\mathrm{OEsoft})^2}((s,\infty);\xi)$ specified by (2.3). Hence, for the soft edge scaling, our sought identity has been established, providing us with a rigorous justification of the identity (1.12).

It remains to establish a similar limit formula for the hard edge. Of the two special pairs of weights (2.18), the second pair near $x=1$ exhibits a general hard edge singularity. Recalling that the ensemble $\mathrm{OE}_N((1-x)^a(1+x)^b)$ with $-1<x<1$ is referred to as the Jacobi orthogonal ensemble $\mathrm{JOE}_N^{a,b}$, the left hand side of (2.17) for the second pair in (2.18) refers to the even labelled eigenvalues counted from the right in the ensemble $\mathrm{JOE}_N^{(a-1)/2,0} \cup \mathrm{JOE}_N^{(a-1)/2,0}$. Analogous to (2.21) the n-point correlation function for the odd labelled eigenvalues of this ensemble is given by the structured formula [**FR04**, eq. (2.16)]

$$
\rho_{(n)}^{\mathrm{odd}(\mathrm{JOE}_N^{(a-1)/2,0})^2}(x_1,\dots,x_n) = \det\Big[-\frac{\partial}{\partial x_j}(1-x_j) \int_{-1}^{x_k} (1-x_j)\tilde{K}_N(x_j,u)\, du \Big]_{j,k=1,\dots,n}. \tag{2.25}
$$

Here $\tilde{K}_N(x,y) = K_N(x,y)/((1-x)(1-y))^{1/2}$ where, with $P^{(a,b)}(x)$ denoting the Jacobi polynomials, $K_N(x,y)$ refers to (1.6) with $g_2(x) = (1-x)^a$ and $p_j(x) = P_j^{(a,0)}(x)$.

In (2.25) the integration variable u is not confined to the neighbourhood of the hard edge. To avoid this potential problem, we make use of the identity [**FR04**, eq. (3.69)]

$$
(1-x)\int_{-1}^1 \tilde{K}_N(x,u)\, du = -2 \int_x^1 \tilde{K}_N(-1,u)\, du \tag{2.26}
$$

where we regard $\tilde{K}_N(-1,u)$ specified by the final form in (1.6), supplemented by the special value $P_j^{(a,0)}(-1) = (-1)^j$.

The hard edge scaling limit in the neighbourhood of $x = 1$ requires replacing the coordinates $x \mapsto 1 - X/2N^2$ then taking $N \to \infty$. To analyze the latter limit, we make use of the uniform asymptotic expansion [**Sze75**]

$$\begin{aligned} &\Big(\sin\frac{\theta}{2}\Big)^a\Big(\cos\frac{\theta}{2}\Big)^b P_n^{(a,b)}(\cos\theta) \\ &\quad = n^{-a}\frac{\Gamma(n+a+1)}{n!}\sqrt{\frac{\theta}{\sin\theta}}J_a((n+(a+b+1)/2)\theta) + \theta^{a+2}O(n^a) \end{aligned} \tag{2.27}$$

valid for $0 < \theta < c/n$, $(c > 0)$, which furthermore remains valid upon differentiation with respect to θ. With use made too of (2.26), this gives

$$(1-x)\int_{-1}^{y}\tilde{K}_N(x,u)\,du\Big|_{\substack{x=1-X/2N^2\\ y=1-Y/2N^2}} = -X^{1/2}\int_Y^\infty v^{-1/2}K^{\rm hard}(X,v)\,dv + O\Big(\frac{1}{N}\Big)O(1), \tag{2.28}$$

which furthermore remains valid upon differentiation, with the dependance on X, Y again $O(1)$. Using the integral form of the kernel (1.16),

$$K^{\rm hard}(x,y) = \frac{1}{4}\int_0^1 J_a(\sqrt{xt})J_a(\sqrt{yt})\,dt$$

we have that

$$\frac{\partial}{\partial X}X^{1/2}\int_Y^\infty v^{-1/2}K^{\rm hard}(X,v)\,dv = K^{\rm hard}(X,y) + \frac{J_a(\sqrt{X})}{2\sqrt{Y}}\int_{\sqrt{Y}}^\infty J_a(t)\,dt.$$

Substituting this in (2.28), substituting the result in (2.25), and recalling (2.6) we conclude

$$\begin{aligned} &\Big(\frac{1}{2N^2}\Big)^n \rho_{(n)}^{{\rm odd}({\rm JOE}_N^{(a-1)/2,0})^2}(1-x_1/2N^2,\dots,1-x_n/2N^2) \\ &\quad = \rho_{(n)}^{{\rm odd}({\rm OEhard})^2}(x_1,\dots,x_n;(a-1)/2) + O\Big(\frac{1}{N}\Big)O(1) \end{aligned} \tag{2.29}$$

where $\rho_{(n)}^{{\rm odd}({\rm OEhard})^2}$ is specified by (2.6). From this uniform asymptotic expansion it is immediate that the analogue of (2.12) holds true, and again by appealing to [**Sos02**] the latter is sufficient for the validity of the limit formula

$$\lim_{N\to\infty} E_N^{{\rm odd}({\rm JOE}_N^{(a-1)/2,0})^2}((1-s/2N^2,1);\xi) = E^{{\rm odd}({\rm OEhard})^2}((0,s);\xi;(a-1)/2), \tag{2.30}$$

with $E^{{\rm odd}({\rm OEhard})^2}$ specified by (2.5). This is the sought limit formula for the hard edge scaling, complementing (2.24) for the soft edge scaling, and providing us with a rigorous justification of (1.13).

References

[AvM95] M. Adler and P. van Moerbeke, *Matrix integrals, Toda symmetries, Virasora constraints and orthogonal polynomials*, Duke Math. Journal **80** (1995), 863–911.

[BD02] A. Borodin and P. Deift, *Fredholm determinants, Jimbo-Miwa-Ueno tau-functions and representation theory*, Commun. Pur. Appl. Math. **55** (2002), 1160–1230.

[BDJ99] J. Baik, P. Deift, and K. Johansson, *On the distribution of the length of the longest increasing subsequence of random permutations*, J. Amer. Math. Soc. **12** (1999), 1119–1178.

[BF03] A. Borodin and P.J. Forrester, *Increasing subsequences and the hard-to-soft transition in matrix ensembles*, J. Phys. A **36** (2003), 2963–2981.

[BR01] J. Baik and E.M. Rains, *Algebraic aspects of increasing subsequences*, Duke Math. J. **109** (2001), 1–65.

[CDG06] O. Costin, P. Deift, and D. Gioev, *On the proof of the universality for orthogonal and symplectic ensembles in random matrix theory*, math-ph/0610063, 2006.

[Dei99] P. Deift, *Orthgonal polynomials and random matrices: A Riemann-Hilbert approach*, Courant Institute of Mathematical Sciences, New York University, 1999.

[DG04] P. Deift and D. Gioev, *Universality in random matrix theoory for orthogonal and symplectic ensembles*, math-ph/0411075, 2004.

[DG06] ———, *Universality at the edge of the spectrum for unitary, orthogonal and symplectic ensembles of random matrices*, math-ph/0507023, 2006.

[DGKV06] P. Deift, D. Gioev, T. Kriecherbauer, and M. Vanlessen, *Universality for orthogonal and symplectic Laguerre type ensembles*, math-ph/0612007, 2006.

[Die05] M. Dieng, *Distribution functions for edge eigenvalues in orthogonal and symplectic ensembles: Painlevé representations*, Int. Math. Res. Notices (2005), 2263–2287.

[Dys72] F.J. Dyson, *Quaternion determinants*, Helv. Phys. Acta **45** (1972), 289–302.

[For] P.J. Forrester, *Log-gases and Random Matrices*, www.ms.unimelb.edu.au/˜matpjf/matpjf.html.

[For93] ———, *The spectrum edge of random matrix ensembles*, Nucl. Phys. B **402** (1993), 709–728.

[For00] ———, *Painlevé transcendent evaluation of the scaled distribution of the smallest eigenvalue in the Laguerre orthogonal and symplectic ensembles*, nlin.SI/0005064, 2000.

[For06] ———, *Hard and soft edge spacing distributions for random matrix ensembles with orthogonal and symplectic symmetry*, Nonlinearity **19** (2006), 2989–3002.

[FR01] P.J. Forrester and E.M. Rains, *Inter-relationships between orthogonal, unitary and symplectic matrix ensembles*, Random matrix models and their applications (P.M. Bleher and A.R. Its, eds.), Mathematical Sciences Research Institute Publications, vol. 40, Cambridge University Press, United Kingdom, 2001, pp. 171–208.

[FR04] ———, *Correlations for superpositions and decimations of Laguerre and Jacobi orthogonal matrix ensembles with a parameter*, Probab. Theory Relat. Fields **130** (2004), 518–576.

[FS05] P.L. Ferrari and H. Spohn, *A determinantal formula for the GOE Tracy Widom distribution*, J. Phys. A **38** (2005), L557–L561.

[Joh00] K. Johansson, *Shape fluctuations and random matrices*, Commun. Math. Phys. **209** (2000), 437–476.

[Joh01] I.M. Johnstone, *On the distribution of the largest principal component*, Ann. Math. Stat. **29** (2001), 295–327.

[Olv74] F. Olver, *Asymptotics and special functions*, Academic Press, London, 1974.

[Por65] C.E. Porter, *Statistical theories of spectra: fluctuations*, Academic Press, New York, 1965.

[Sas05] T. Sasamoto, *Spatial correlations of the 1D KPZ surface on a flat substrate*, J. Phys. A **38** (2005), L549–L556.

[Sos02] A. Soshnikov, *A note on the universality of the distribution of the largest eigenvalues in certain sample covariance matrices*, J. Stat. Phys. **108** (2002), 1033–1056.

[Sze75] G. Szegö, *Orthogonal polynomials*, 4th ed., American Mathematical Society, Providence R.I., 1975.

[TW94] C.A. Tracy and H. Widom, *Level-spacing distributions and the Airy kernel*, Commun. Math. Phys. **159** (1994), 151–174.

[TW96] ———, *On orthogonal and symplectic matrix ensembles*, Commun. Math. Phys. **177** (1996), 727–754.

[WW65] E.T. Whittaker and G.N. Watson, *A course of modern analysis*, 2nd ed., CUP, Cambridge, 1965.

Current address: Department of Mathematics and Statistics, University of Melbourne, Victoria 3010, Australia

E-mail address: `p.forrester@ms.unimelb.edu.au`